Smart Internz

Detecting Phishing websites using IBM watson studio

Applied Data Science Final project

Submitted by -

Aagusthya Shanker - 20BCE10140 - aagusthya.shanker2020@vitbhopal.ac.in

Arvind Pratap Singh - 20BCE10896 - arvind.pratap2020@vitbhopal.ac.in

Prakhar Gupta - 20BCE10542 - prakhar.gupta2020@vitbhopal.ac.in

1. Introduction :

- The proliferation of the internet and the increasing reliance on online platforms for various activities have given rise to a significant threat known as phishing. Phishing refers to the deceptive practice of fraudsters creating fake websites that mimic legitimate ones in order to trick users into revealing sensitive information such as passwords, credit card details, or personal data. This fraudulent activity poses a severe risk to individuals, organizations, and the overall security of the digital landscape.

- Detecting phishing websites is a critical challenge as attackers constantly adapt their techniques to bypass traditional security measures. To address this issue, this project focuses on leveraging website links as a means to identify and classify potential phishing websites accurately. By analyzing the characteristics and patterns of URLs, we aim to develop a robust detection system that can quickly and accurately identify fraudulent websites, thus empowering users to protect themselves against phishing attacks.

1.1 : Motivation :

- In today's digital age, phishing attacks have become a major concern for individuals and organizations alike. Phishing attacks can result in devastating consequences such as data breaches, financial losses, identity theft, and reputational damage. As such, studying phishing attacks can have far-reaching benefits in terms of identifying vulnerabilities, developing effective countermeasures, and raising awareness among the general public.

- Phishing attacks have become one of the most prevalent forms of cybercrime in recent years, affecting millions of people worldwide. Phishing attacks are designed to trick individuals into divulging sensitive personal or financial information, such as passwords, credit card numbers, and social security numbers. The rise in phishing attacks has created a pressing need for research into how these attacks work, why people fall victim to them, and what can be done to prevent them.

- We, the students interested in Cybersecurity,Artificial Intelligence,Data Science,Core Computer Science believe that studying phishing attacks would be an ideal project to undertake. Not only is this a crucial topic in the field of cybersecurity, but it is also a relevant issue that affects people from all walks of life. By researching and analyzing phishing attacks, we can help shed light on the methods used by attackers, as well as how individuals and organizations can better protect themselves against these threats.

- One of the reasons that phishing attacks are so successful is that they are often very convincing. Attackers use a variety of techniques, such as mimicking legitimate websites or creating fake email addresses that look like those of legitimate organizations, to trick users into revealing sensitive information. In order to better understand how these attacks work, it is necessary to examine the tactics used by attackers, as well as the vulnerabilities that allow individuals and organizations to fall prey to them.

- Furthermore, the consequences of falling victim to a phishing attack can be severe, both for individuals and organizations. In addition to the risk of financial loss or identity theft, phishing attacks can also lead to reputational damage for companies that are targeted. By studying the various types of phishing attacks and their potential impact, we can help develop better strategies to prevent and mitigate the damage caused by these attacks.

- Another reason why researching phishing attacks is important is that they are constantly evolving. Attackers are always looking for new ways to exploit vulnerabilities in software, hardware, and human behavior. By staying up-to-date on the latest phishing attack techniques, we can help organizations anticipate and prepare for new threats as they emerge.

- Moreover, as technology continues to advance, phishing attacks are likely to become even more sophisticated and difficult to detect. As a result, it is important to develop new tools and techniques to combat these threats. By researching phishing attacks, we can help develop and test new strategies for protecting individuals and organizations against these attacks.

- Finally, studying phishing attacks can help raise awareness of the issue among the general public. Many people are unaware of the risks posed by phishing attacks, and may not know how to protect themselves against them. By conducting research and sharing my findings with others, we can help raise awareness of the issue and encourage individuals to take steps to protect themselves against these threats.

- In conclusion, studying phishing attacks is a valuable and important project for anyone interested in cybersecurity. By researching and analyzing these attacks, I can help develop better strategies for protecting individuals and organizations against these threats. We can also help raise awareness of the issue and encourage individuals to take steps to protect themselves. With the increasing prevalence and severity of phishing attacks, there has never been a more pressing need for research in this field.

- Our project “Detecting Phishing websites using ml models” helps people to identify whether a website is legitimate or illegitimate so that they are aware whether they are being phished or not.

2. Literature survey :

Note: The focus of our discussion will primarily be on India and not other countries, as we have chosen India as the scope for our pilot project. This should not be confused with the scenarios of other countries.

2.1: Existing work and its limitations :

- In India, several organizations and initiatives are working to mitigate the threat of phishing. One such initiative is the Indian Computer Emergency Response Team (CERT-In), which operates under the Ministry of Electronics and Information Technology. CERT-In is responsible for dealing with cybersecurity incidents, including phishing attacks. CERT-In provides a number of services to individuals and organizations in India to help them improve their cybersecurity posture and defend against phishing attacks. These services include alerts and advisories on emerging threats, training and awareness programs, and free tools and services to help individuals and organizations secure their systems and networks.

- The Reserve Bank of India (RBI) has also taken several steps to mitigate the threat of phishing. The RBI has issued guidelines to banks and financial institutions on how to prevent and respond to phishing attacks. These guidelines include measures such as implementing multi-factor authentication, conducting regular security awareness training for employees, and implementing strict access controls for sensitive information. The RBI has also established a Cyber Security and Information Technology Examination (CSITE) cell to oversee the cybersecurity practices of banks and financial institutions in India.

- The Data Security Council of India (DSCI) is another organization that is working to mitigate the threat of phishing in India. The DSCI is a not-for-profit organization that was set up by NASSCOM, a trade association for the Indian IT industry. The DSCI provides training, certification, and best practice guidance to help Indian organizations improve their cybersecurity posture. The DSCI also works closely with the government to develop policies and regulations related to cybersecurity.

- In addition to these initiatives, many Indian organizations are taking steps to improve their own cybersecurity posture and protect against phishing attacks. For example, some organizations are implementing email filtering solutions to prevent phishing emails from reaching employees. Others are using phishing simulation exercises to train employees to recognize and respond to phishing attacks. Still others are implementing stricter access controls for sensitive information, including the use of multi-factor authentication and role-based access controls.

- Despite these efforts, phishing attacks continue to be a significant problem in India. One reason for this is the increasing sophistication of phishing attacks. Attackers are using advanced techniques such as spear phishing, which involves targeting specific individuals with personalized messages that appear to be from a trusted source. Another reason is the lack of cybersecurity awareness among the general public. Many individuals in India are not aware of the risks associated with using technology, and may not know how to recognize and respond to phishing attacks.

- To address these challenges, there is a need for continued investment in cybersecurity initiatives in India. This includes investment in education and awareness programs to help individuals and organizations understand the risks associated with using technology, as well as investment in technologies such as artificial intelligence and machine learning to help detect and prevent phishing attacks.

- In conclusion, phishing is a significant threat to individuals and organizations in India, and requires a concerted effort from government, industry, and individuals to address. The initiatives and organizations discussed in this literature review are important steps in the right direction, but more needs to be done to mitigate the threat of phishing in India. By continuing to invest in cybersecurity initiatives and working together, we can help to make India a safer and more secure place for individuals and organizations to use technology.

2.2 : Proposed Work :

- **To create a phishing website which consists of a phishing url checker and general information related to phishing. This Phishing URL checker will help people find whether a website is used for phishing or its a legitimate website.**

3. Theoretical Analysis :

3.1 : Block Diagram :

3.2 : Hardware / Software Designing :

- **Working environment :**
  - Beautiful soup4 == 4.11.2
  - bs4==0.0.1
  - Html5lib == 1.1
  - Dnspython == 2.3.0
  - Pymongo == 4.3.3
  - Scikit-learn == 1.2.2
  - Base64 == pre-installed
- **Browsers :**
  - Google Chrome
  - Safari
  - Mozilla Firefox
  - Microsoft Edge
  - Amazon Silk
- **Operating Systems :**
  - Windows 7 ,10 ,11
  - Linux
  - MacOS
  - Android
  - IOS
- **Hardware Requirements :**
  - Processor = 1 gigahertz (GHz) or faster processor
  - Storage Space = 1 GB >
  - RAM = 2 GB DDR4 >

- Hard disk = Any (HDD or SDD)
- Graphic card = Integrated 2 GB 64 bit >

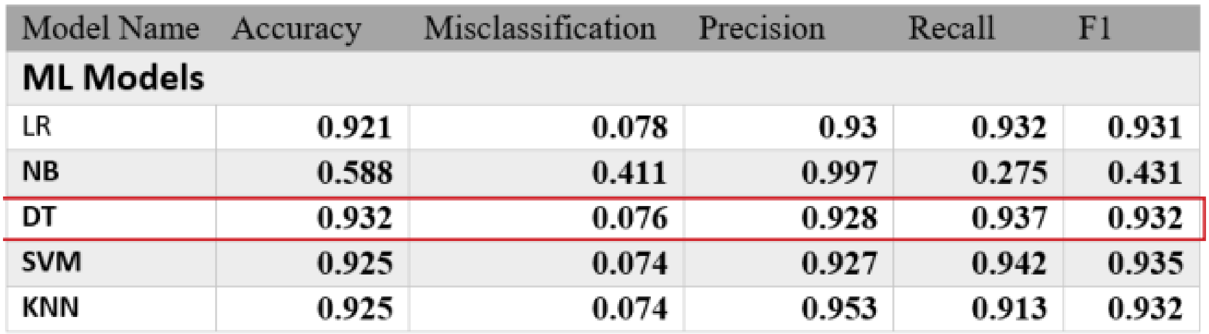
4. Experimental Investigations :

- We have provided infographs of website performance that helps you in understanding the website and predict the website's performance.

- Bit by Bit analysis gives you information about the specific inputs the website is entertaining, the demographic details the website has etc

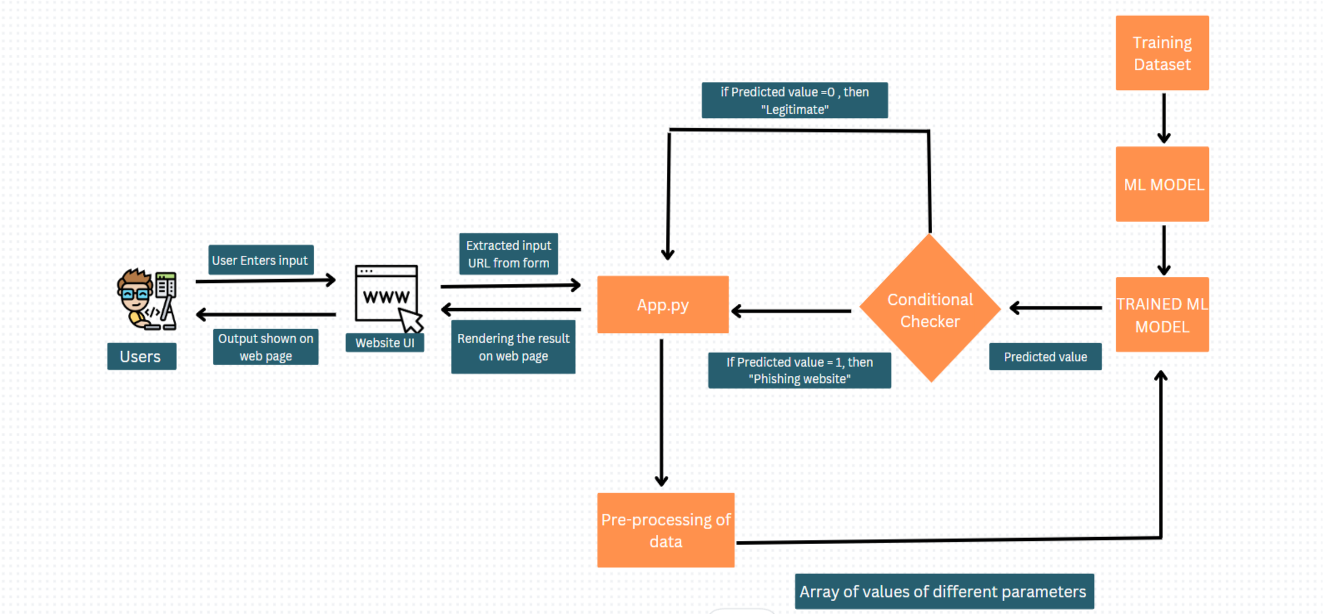
- This analysis is done to re-iterate the fact that investing in data analysis will assist organizations and individuals in saving oneself from phishing and better decision-making.

- Our project done on limited data helps the viewers and organizations to fully understand the potential risks of phishing and its drawbacks.
- In this project we have used Decision trees classification algorithm as it out-performed all the other ml models in accuracy metric.



Model Name	Accuracy	Misclassification	Precision	Recall	F1
ML Models					
LR	0.921	0.078	0.93	0.932	0.931
NB	0.588	0.411	0.997	0.275	0.431
DT	0.932	0.076	0.928	0.937	0.932
SVM	0.925	0.074	0.927	0.942	0.935
KNN	0.925	0.074	0.953	0.913	0.932

5. Flowchart :

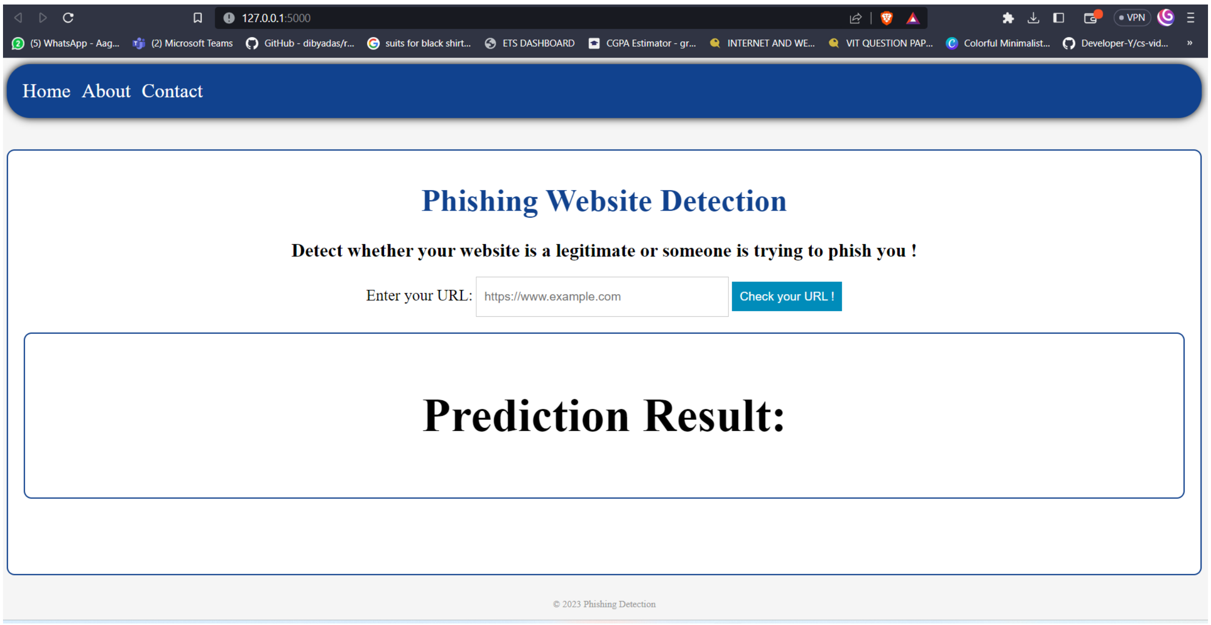

Training Dataset
ML MODEL
if Predicted value =0 , then "Legitimate"
User Enters input
Extracted input URL from form
WWW
Users
Output shown on web page
Website UI
Rendering the result on web page
App.py
Conditional Checker
TRAINED ML MODEL
If Predicted value = 1, then "Phishing website"
Predicted value
Pre-processing of data
Array of values of different parameters

6. Result:

The data product (website) created through this project is intended to construct a onestop webapp for all the Phishing website data and analysis and also sue in real-time as the Phishing is current boom in the Security Industry. This data is collected from various authentic websites. Later the data is arranged, organized in such a way that it fills the gaps in information. The pages created in the webapp give complete information and analysis of Phishing data be it Socket Number, SSL ID, Website URL and Prediction of the authenticity of the matches to a little extent, etc., In this project we aim to impart the ability to get required information from a single site rather than searching multiple sites. Specifically, we will aim to go beyond information retrieval to do reasoning over the multimodal dataset for easy access of required information. Site visitors are at liberty to use one or more of these datasets to interpret, predict, draw intelligence of any sort from the dataset provided.

Here are the output screenshots of our website :

Home page :



Phishing Website Detection

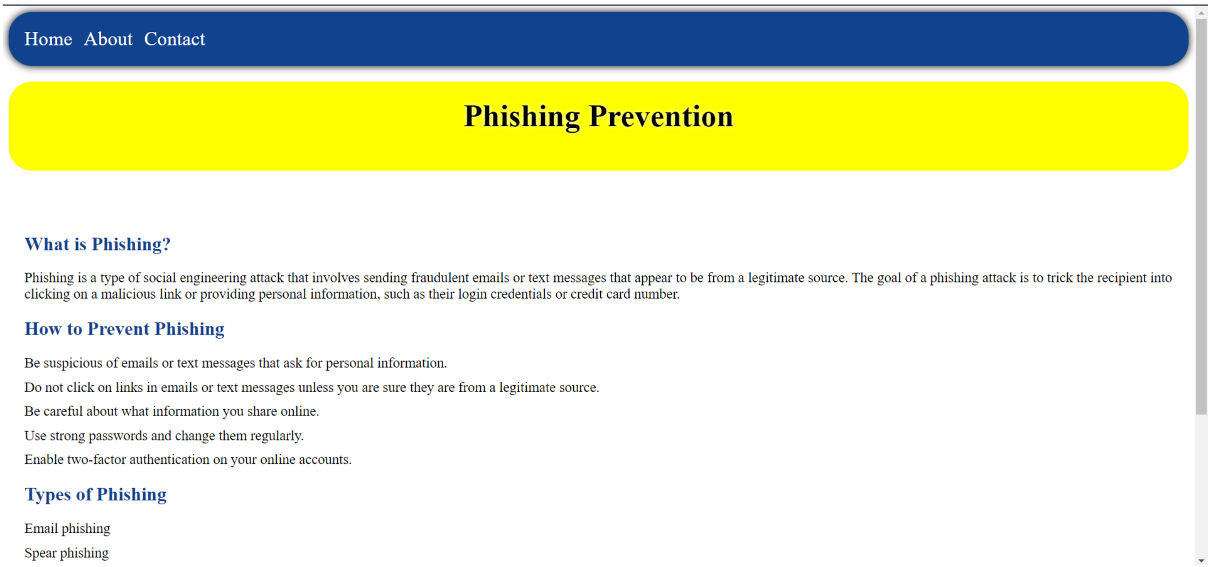
Detect whether your website is a legitimate or someone is trying to phish you !

Enter your URL: https://www.example.com

Prediction Result:

© 2023 Phishing Detection

About page:



Phishing Prevention

What is Phishing?

Phishing is a type of social engineering attack that involves sending fraudulent emails or text messages that appear to be from a legitimate source. The goal of a phishing attack is to trick the recipient into clicking on a malicious link or providing personal information, such as their login credentials or credit card number.

How to Prevent Phishing

Be suspicious of emails or text messages that ask for personal information.

Do not click on links in emails or text messages unless you are sure they are from a legitimate source.

Be careful about what information you share online.

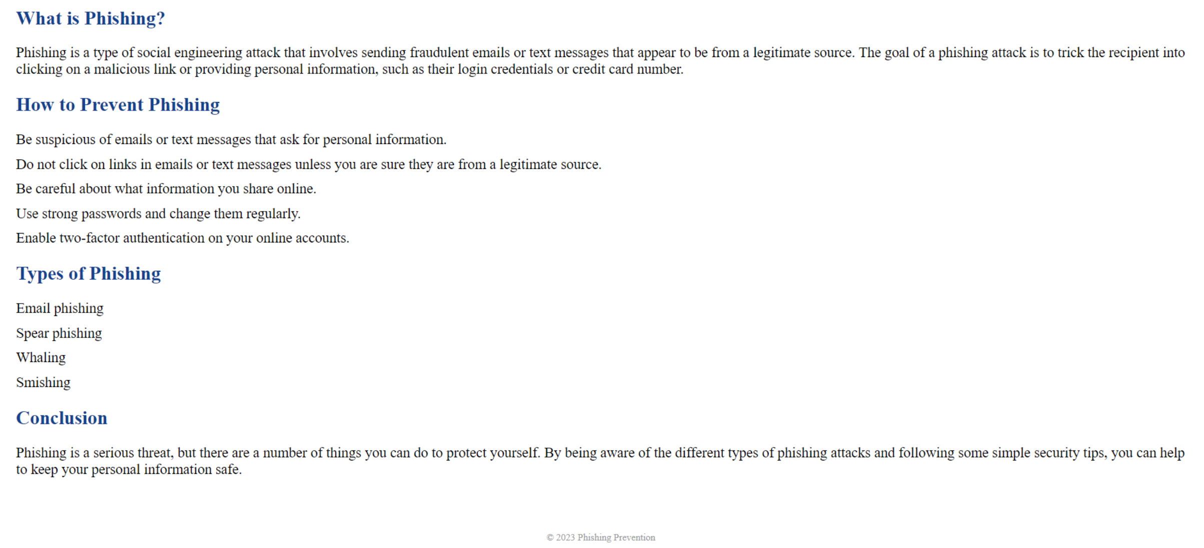
Use strong passwords and change them regularly.

Enable two-factor authentication on your online accounts.

Types of Phishing

Email phishing

Spear phishing



What is Phishing?

Phishing is a type of social engineering attack that involves sending fraudulent emails or text messages that appear to be from a legitimate source. The goal of a phishing attack is to trick the recipient into clicking on a malicious link or providing personal information, such as their login credentials or credit card number.

How to Prevent Phishing

Be suspicious of emails or text messages that ask for personal information.

Do not click on links in emails or text messages unless you are sure they are from a legitimate source.

Be careful about what information you share online.

Use strong passwords and change them regularly.

Enable two-factor authentication on your online accounts.

Types of Phishing

Email phishing

Spear phishing

Whaling

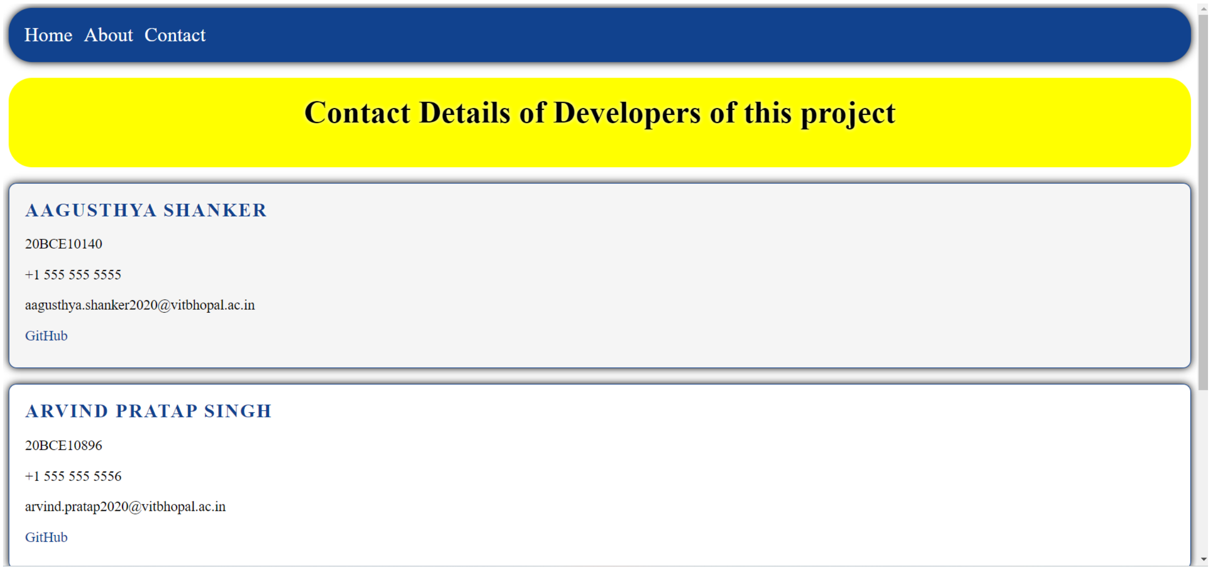
Smishing

Conclusion

Phishing is a serious threat, but there are a number of things you can do to protect yourself. By being aware of the different types of phishing attacks and following some simple security tips, you can help to keep your personal information safe.

© 2023 Phishing Prevention

Contact page :



Contact Details of Developers of this project

AAGUSTHYA SHANKER

20BCE10140

+1 555 555 5555

aagusthya.shanker2020@vitbhopal.ac.in

GitHub

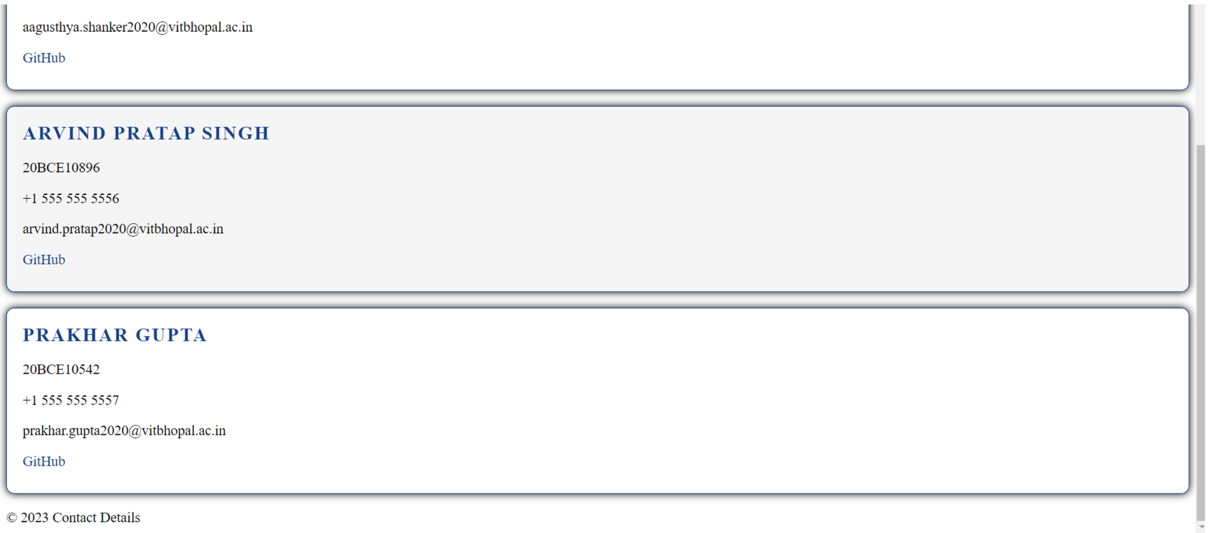
ARVIND PRATAP SINGH

20BCE10896

+1 555 555 5556

arvind.pratap2020@vitbhopal.ac.in

GitHub



aagusthya.shanker2020@vitbhopal.ac.in

GitHub

ARVIND PRATAP SINGH

20BCE10896

+1 555 555 5556

arvind.pratap2020@vitbhopal.ac.in

GitHub

PRAKHAR GUPTA

20BCE10542

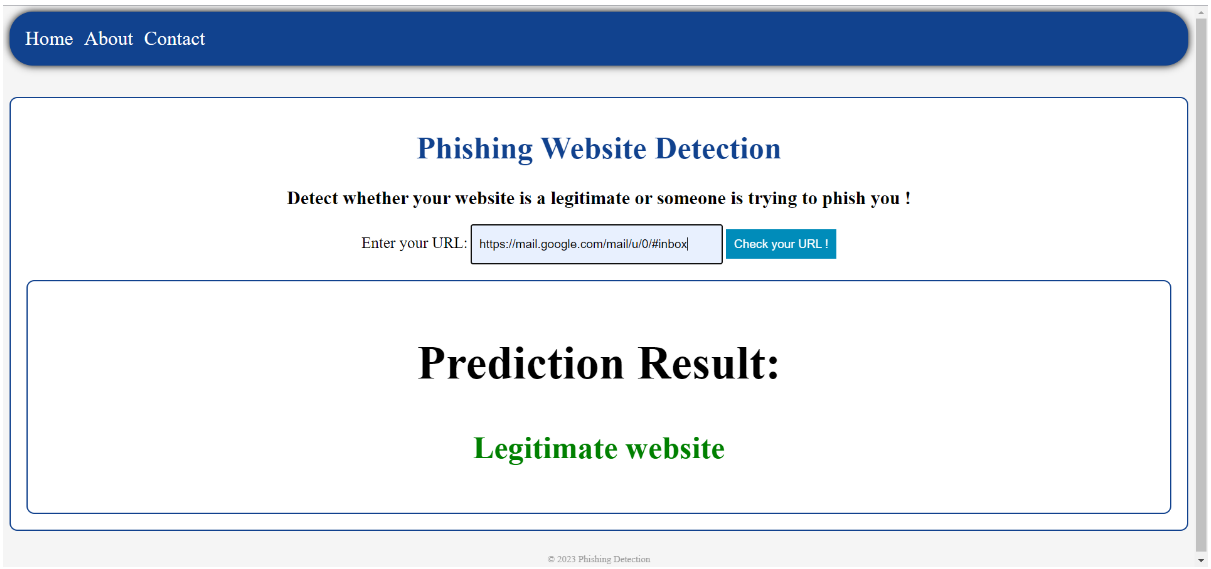
+1 555 555 5557

prakhar.gupta2020@vitbhopal.ac.in

GitHub

© 2023 Contact Details

Prompting a legitimate website URL on URL checker :



Phishing Website Detection

Detect whether your website is a legitimate or someone is trying to phish you !

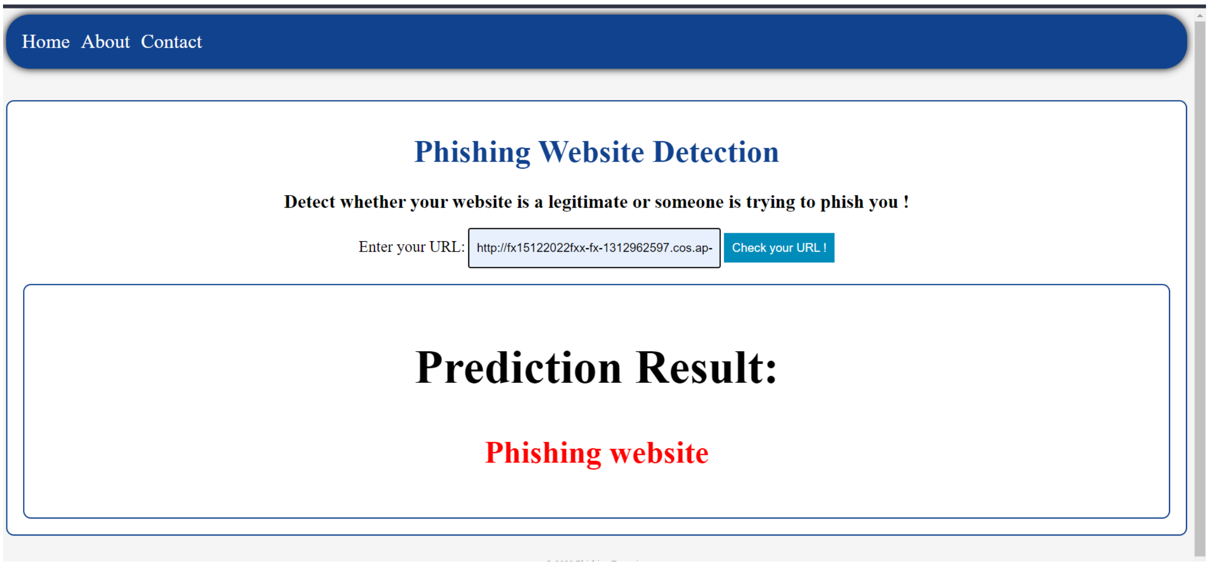
Enter your URL: https://mail.google.com/mail/u/0/#inbox

Prediction Result:

Legitimate website

© 2023 Phishing Detection

Prompting a phishing website URL on URL checker :



Phishing Website Detection

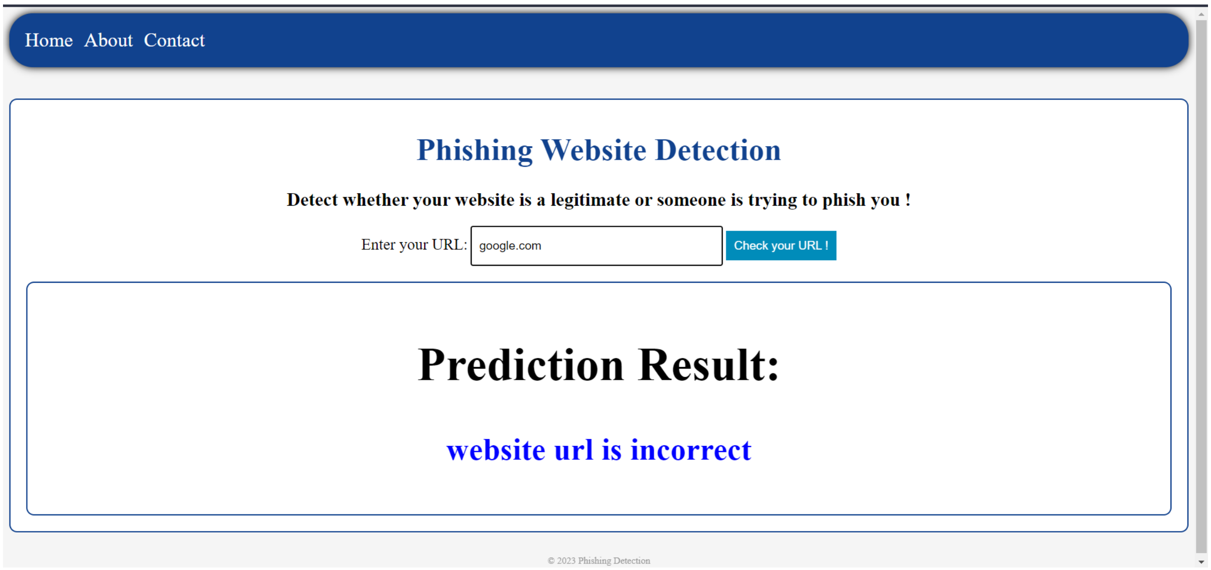
Detect whether your website is a legitimate or someone is trying to phish you !

Enter your URL: http://fx15122022fxx-fx-1312962597.cos.ap-

Prediction Result:

Phishing website

Prompting an incorrect URL on URL checker :



Phishing Website Detection

Detect whether your website is a legitimate or someone is trying to phish you !

Enter your URL: google.com

Prediction Result:

website url is incorrect

© 2023 Phishing Detection

7.Advantages and disadvantages :

- The detection of phishing websites using website links is crucial in safeguarding individuals and organizations from falling victim to online scams and cyberattacks. By developing an accurate and efficient detection system, this project aims to:

a. Enhance user awareness: Empower internet users with the knowledge and tools to identify potential phishing websites, enabling them to make informed decisions while navigating the online landscape.

b. Mitigate financial losses: Minimize the financial impact caused by phishing attacks by proactively identifying and blocking fraudulent websites, thereby protecting individuals and organizations from financial fraud.

c. Safeguard data security: Preserve the privacy and security of sensitive information by preventing unauthorized access and disclosure through phishing attempts.

d. Contribute to research and innovation: Advance the field of cybersecurity by proposing novel techniques and insights into the detection of phishing websites, paving the way for future advancements in online security.

8. Applications :

- The Phishing Website analysis web app is constructed to reiterate the importance of data analytics in a day-to day setting. Some Website may provide data in various forms, the information provided by websites is not comprehensive and has information gaps. In order to get the required information, the viewer has to visit many sites and comprehend the information. Our site removes all those hurdles and provides information and analysis with a single click.

- Data analysis helps organizatins to know their customer and paves the way for innovative solutions, hyper targeted strategies, and personalized marketing campaigns. Utilizing analytics will unlock a whole range of competitive advantages. Data analysis helps in better decision-making, more accessible analytics, automation, predictive modeling.

9. Conclusion :

- Individuals and Organizations can become truly customer-centric by using data analysis. Delivering a personalized experience is considered the standard that most marketing teams strive for. With the models we have incorporated in preparing the data product, the organizations can provide tailored interactions at scale. Consumers tend to revisit the site if they get personalized views to those perceived as being generic.

- Moreover, we also made it open source so that it can be scaled up, consolidated, cost-effective, integrated management and so on.

10. Future scope :

- In future, we can work upon the user-interface of the website to make it more attractive and easy to use. Also, in this project we have only used machine learning models for classification, however there are many other learning techniques which might give more accurate results than the current model in use.

11. Bibliography :

1. **ANTI-PHISHING WORKING GROUP: HTTPS://WWW.ANTIPHISHING.ORG/**
2. **PHISHING.ORG: HTTPS://WWW.PHISHING.ORG/ FEDERAL TRADE**
3. **COMMISSION - CONSUMER INFORMATION ON PHISHING: HTTPS://WWW.CONSUMER.FTC.GOV/ARTICLES**
4. **HOW-RECOGNIZE-AND-AVOIDPHISHING-SCAMS UNITED STATES**
5. **COMPUTER EMERGENCY READINESS TEAM - PHISHING: HTTPS://USCERT.CISA.GOV/NCAS/TIPS/ST04-014**